

**Project Name:** MASK (Mark, Achyut, Sam, Kiran)

**Project Scope:** Extract, Transform and Load Transportation Data

**Data Source:** <https://www.kaggle.com/citylines/city-lines>

**Date:** 20-Aug-2019

## Overview

### 1. Project Background and Description

**Background:** This project is to identify how lines have changed over time based on historical records.

**Contributors:** Mark, Achyut, Sam, Kiran

**Description:** This dataset, originally shared, includes transportation line data from a number of cities from around the world including London, Berlin, Mexico City, Barcelona, Washington D.C., and others, covering many thousands of kilometers of lines.

It consists of below tables available in its raw form.

Data Sources	
cities.csv	334 x 7
lines.csv	1343 x 7
station_lines.csv	16.2k x 6
stations.csv	15.8k x 7
systems.csv	488 x 3
track_lines.csv	9757 x 6
tracks.csv	9271 x 7

## 2. Project Scope

This project is to collect data related to transportation systems across the world and transform the data into desired format and insert it into a database.

## 3. High-Level Requirements

Collect Data from any external source

Review the dataset

Cleanup the junk data

Combine the datasets to group the data by Country and City

Create a finalized database for significant future insights

## 4. Deliverables

All the datasets merged in a Postgres Database.

## 5. Implementation Plan

Create solutions for transportation by combing through our dataset in order to find inadequacies.